Fermi National Accelerator Laboratory

FERMILAB-Conf-96/015-E
SDSS

The Sloan Digital Sky Survey

Thomas Nash

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

January 1996

Proceedings of the *IInd Rencontres Du Vietnam, The Sun and Beyond*,
Ho Chi Minh City (Saigon), Vietnam, October 21-28, 1995

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CHO3000 with the United States Department of Energy

THE SLOAN DIGITAL SKY SURVEY

Thomas Nash*
Fermi National Accelerator Laboratory
P.O.Box 500
Batavia, IL 60510 USA
nash@fnal.gov
(*for the SDSS Collaboration)

Systematic large area surveys of objects in the night sky were among the earliest scientific undertakings, dating back at least to Ptolemy. The persistent relevance of data from maps of the cosmos is unmatched for any other class of scientific data. Photographic plates from the 19th century Carte de Ciel survey are still referenced to measure proper motion and other time variable effects. The Palomar Sky Survey has been the underpinning of much astronomical research for the last four decades.

Photographic plates lack the sensitivity, linearity, and dynamic range of CCD detectors. Photographic surveys are thereby less complete and less unbiased in their inclusion criteria and measured variables. It has only now become possible to build CCD arrays with area comparable to, or larger than, plates. The Sloan Digital Sky Survey (SDSS), which I describe here, will produce a photometric survey of unprecedented scale (one quarter of the sky - π sr - with about 10^8 galaxies and 10^6 quasars) and high precision (<2% accuracy and reproducibility).

In the last decade, digital techniques have been used with great effectiveness to produce three dimensional "red shift" maps of portions of the sky. The survey by Geller and Huchra[1)] discovered structure measuring 150 h^{-1} Mpc by 50 h^{-1} Mpc, the "Great Wall", and established a strong scientific interest in very large scale structure as a tracer of physics in the early universe and a constraint on the cosmology that created the universe we see today. Approximately 18,000 galaxy red shifts were measured over a dimension which just exceeded the scale of the Great Wall.

The SDSS is aiming to measure the red shift of some 1 million galaxies and 100,000 QSOs. This approaches a two order of magnitude increase in statistics over the seminal work of Geller and Huchra. The survey will span a volume some fifty times larger, large enough to determine the frequency of occurrence of Great Wall sized structures and to measure the power spectrum out to scales that overlap the COBE measurements of structure in the cosmic microwave background (CMB) (Figure 1). For the first time, this will permit direct comparison of structure (and the power spectrum) of present day matter with that of the driving force for this structure, the gravitational potential on the scale of the horizon at the time of matter-radiation equality in the early universe, as reflected by the CMB. Most important, this survey will improve on the red shift candidate selection of previous surveys which selected objects for red shift measurements from smaller photographic plate surveys, based on sometimes poorly defined criteria. The SDSS will select candidates with unparalleled uniformity from the extraordinary digital photometric catalogue which will be collected concurrently on the same dedicated 2.5 meter telescope.

Figure 1. Anticipated SDSS measurement of power spectrum (solid points). IRAS data (open points) shown for comparison. Scales probed by COBE anisotropies are indicated.

The SDSS project is being built, and will be operated, by a collaboration consisting of the University of Chicago, Fermi National Accelerator Laboratory, Institute for Advanced Study, the Japanese JPG astronomy group, Johns Hopkins University, Princeton University, US Naval Observatory, and the University of Washington. The survey plan calls for a run of 5 years covering half the northern sky. Installation of the telescope, instrumentation, and other systems required for the survey is presently underway at

Apache Point, New Mexico, and a commissioning and scientific test period should commence this year.

Table I is a list of project goals for high level scientific figures of merit. Although these may be relaxed (or exceeded) as a result of experience during commissioning of the complex systems I will describe below, they are the basis for our expectations that this project will reap an enormous scientific yield. The science spans a large range of astronomy and astrophysics interests from basic stellar statistics to galaxy and QSO evolution at early times.

Table I. Preliminary High Level Goals

Subject to change based on actual conditions and experience.

Survey Run Time	5 years
Main survey area	π sr North Galactic Pole region
Deep Photometric Survey	
Uniformity in m(r’)	2% rms over all sky of survey
Magnitude limits	m(u’)=21.9, m(g’&r’)=22.5, m(i’)=22.0, m(z’)=20.4 (5 sigma for stellar objects)
Completeness	95%
Contamination	5%
Image size	1 arcsec
Spectroscopic Survey	
Number of red shifts	galaxies 10^6, QSOs 10^5
Redshift accuracy	$\Delta Z < 30$ km / sec
Bright star exclusion region	<5% sky
Redshift completeness	galaxies 90%, QSOs 72%
Minimum object separation	50 arcsec (one pass, minimum fiber spacing)

Looking at just one general subject area, large-scale structure, the scope of key science topics is impressive: redshift space power spectra and correlation functions, their anisotropy, higher order correlations, topology of structure, angular correlation functions, angular-redshift cross correlations, clustering with photometric redshifts, species dependent clustering, distances to elliptical galaxies, QSO clustering, and QSO absorption-line clustering. One example we cited, the topology of the space distribution of clusters and groups, indicates how the Sloan survey will allow scientific measurements of structures at a level of sophistication well beyond the traditional measurements of power spectra and correlation functions. This approach suggested by Gott *et al.* 2) is shown in

Figure 2. At a particular smoothing length (here $\lambda = 600$ km / sec), the genus of isodensity surfaces is plotted versus the threshold density, ν. Genus is the number of donut like holes minus the number of isolated regions of the surface. Large positive values describe a Swiss cheese like structure; large negative values are meat ball like; and around zero the structure is sponge like. With sufficient statistics and precision, these distributions are sensitive to cosmological model predictions, such as those of non-gaussian primordial fluctuations like cosmic strings. Figure 2a and b compare the status of this information from presently available data with what can be expected from the SDSS data.

Figure 2. Comparison of quantitative measurement of the topology of galaxy distributions. a) Left, results from existing large red shift surveys. b) Right, results expected from the SDSS. Variables explained in text.

Before we return to some other examples of science we expect may come out of the survey, let's look at the complex systems that will bring us this data. The survey is based on the philosophy that the best 20% of seeing time should be dedicated to the photometric survey, from which one can select candidate objects for spectroscopic red shift measurements in the remainder of the observing time. To accomplish the goal of completing the survey in a five year period, we will require efficient operations, some reasonable luck with the weather, and a dedicated telescope with a large and well mapped field of view. The telescope has a 2.5 m primary mirror and a very large (over 1 m) secondary to produce a 3° field at the focal plane with an overall aperture ratio of $f/5$. The mount is altitude-azimuth. The telescope, except for the glass, has been finished and installed at Apache Point Observatory in New Mexico at an elevation of 2800 meters.

As shown in Figure 3, the telescope breaks new ground in terms both of reduced cost and greatly improved control of seeing degradation by air temperature gradients and turbulence in that its enclosure rolls away to a downwind position during observations. The telescope then stands free on a platform extending out over a ridge above the tree tops. There is no telescope dome to induce turbulence and seeing degradation; air currents can move up the ridge in a laminar fashion past the telescope. A temperature controlled wind baffle surrounds the telescope and moves on an independent mount that tracks the telescope motion.

Figure 3. The Sloan Digital Sky Survey Telescope at Apache Point Observatory, NM

The critical technology enabling this *digital* photometric survey of the sky is an array of 2048 x 2048 element CCDs in six columns of five. This array spans approximately 40 x 60 cm of the center of the focal plane. The five chips in each column look through filters defining the five primary bands of the survey (Figure 4). Leading and following the photometric chips are twelve 2048 x 400 chips for astrometric calibration and focus. All the chips meeting our quality requirements are in hand, and this huge camera is being assembled. The camera will operate in “drift scan mode” (time-delay-and integrate, TDI imaging); the telescope will track along great circles at the sidereal rate with the CCDs aligned so that images move along their shift columns. The CCD clock rate is set to shift charge so it remains in spatial sync with the image from which it came. When the charge shifts out of the chip it represents the 55 seconds of integrated light detected for that point in the sky. Because of the gap between the chips, it is necessary to scan each region of the sky a second time with an offset; ~ 20% of the sky will be in overlap regions and imaged twice, permitting calibration checks and searches for variable phenomena.

Figure 4. SDSS Filter Response Curves (with and without atmospheric absorption)

The ambitious goal of maintaining <2% photometric systematic accuracy across π sr in a five year survey will depend on the use of a small robotic monitor telescope located less than 50 meters from the main survey telescope. A set of fundamental photometric standard stars will be observed each night in the regions being observed by the main survey, as nearly simultaneously as possible. The monitor telescope is a 0.6 meter *f*/10 Ritchey Chrétien design with correctors capable of 0.6” images over its field of view on a 2048x2048 CCD. The monitor telescope will determine extinction coefficients on an hourly basis for various airmasses, and it will determine photometric zero points by calibrating a grid of transfer star fields which include standard stars. Photometric solutions will be calculated in real time so that observers on shift may monitor the quality of the night.

Similar monitoring of spectrophotometric standards will be carried out during spectroscopic runs.

The photometric survey will be carried out in the best seeing conditions, approximately 20% of the observing time. The remainder of the time will be devoted to a spectroscopic survey. The spectra of over 600 objects in a region called a “tile” will be taken at one time in exposures of a little less than an hour. An aluminum plate with holes drilled at the anticipated positions of spectroscopic candidate objects will be located at the focal surface. Optical fibers plug into the holes and carry light to one of a pair of twin spectrographs, where the light is divided into two paths by a diachronic beam splitter and carried to a “blue” (3900 - 6000 Å) and a “red” (6000 - 9100 Å) camera and grisms (Figure 5). CCDs will image the spectra, which will then be recorded and available for red shift and other spectral analysis.

Figure 5. A cross-sectional view of one of the two spectrographs.

All galaxies up to a well defined magnitude of g’<19, and all quasars to g’< 20, that are identified from the data of the photometric survey will be targets for the spectroscopic survey. The remaining available spectrographic fiber channels will be used for a selection of stellar objects in the field with a few reserved for searches for serendipitous targets of opportunity. Galaxy identification is based on the spatial extent of their morphology. Quasars will be distinguished from stellar objects based on color-color graphs using the five filter observations of each object in the photometric survey. We are aiming to measure a million galaxy and 100,000 QSO red shifts in the full spectroscopic survey. Beyond this, less precise but still valuable, red shift measurements of possibly 10% of the 100 million galaxies in the photometric survey will be made using the 5 color observations (see Figure 6). Spectroscopic measurements of objects as close as 1 arcmin are possible in one run. The sky will be tiled according to an adaptive algorithm which puts the tiles closer together in dense regions. We can observe many close object pairs separately in the tile overlap regions.

Figure 6. Red shift estimated from photometric colors vs. spectroscopic red shift (from D. Koo & R. Kron, *Ann. Rev. Astron. Astrophys.*,**30,** 613). For the SDSS we expect $\Delta Z < 0.2$.

The survey will produce an enormous amount of data, the total raw data will exceed 20 Terabytes. This alone would require sophisticated and extensive data acquisition, data reduction, and analysis systems. Even more than this, the essential philosophy and strategy of the survey, spectroscopic targets selected with little bias from a large homogeneous photometric survey running on good seeing nights, requires an expeditious and efficient data handling system in order that hole positions for the spectroscopic plug plates may be calculated and drilled in time for the next lunation. The ambitious goals for the scope of the survey in a defined time period of five years further requires a high level of efficiency in monitoring the quality of observations and the proper functioning of systems, so that problems are identified and repaired with little down time.

The data acquisition and reduction system is built along the model of modern high energy physics experiments using electronics in the VME standard, read out in parallel under the supervision of a Silicon Graphics computer. Photometric data is written in ten parallel streams (the six photometric columns of 5 CCDs and four astrometric/focus sets of CCDs) to DLT tape drives. A similar VME data readout system is used for the spectrographic system. One copy of data will be retained at the observatory and a second sent by overnight express to the main computer facility of the project in the Feynman Computer Center at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, IL. There the data will be reduced through photometric and astrometric, or spectroscopic, “pipeline” software developed expressly for the project by scientific members of the project institutions. This software takes advantage, where appropriate, of existing astronomy and other scientific software. The data reduction will be carried out on a DEC Alpha 8200 computer, with a performance capacity of about 3500 mips.

The computing systems were built by computer professionals and engineers at Fermilab, who created the skeleton, and standards, in which is embedded the science sensitive software, developed by scientists. Much use is made of the interactive language, TCL, which is the foundation on which many testing and analysis tools sit in a way that allows rapid prototyping and testing of new ideas. Object oriented data bases (based on the product Objectivity) are used in two places. The operational data base is used to run the survey, to keep track of sky tiling, to identify target objects for hole drilling, etc. It maintains the results of the data reduction, identified objects, calibration constants, atlas images, etc. The analysis data base also stores this reduced (calibrated, strip matched, color merged, etc.) output of the survey in a form designed to make scientific analysis easy and reduce the latency for producing scientific results.

Table II shows the data reduction steps. The project is committed to releasing the data to the public. The first release will be of the first two years of data at a time two years after this data has been taken. The final release will be two years after completion of the five year survey. Data will be available at all levels, including single CDROMs (or the technological equivalent appropriate at the time of release) for amateurs and schools. Atlas images are small regions of sky around each identified object. They are intended to allow researchers to go back and reanalyze the parameters of an object, for example to check biases and calibrations. Rarely will it be necessary for any but data reduction experts to go back to the raw data.

Table II Hierarchy of Data

Content	Size
Raw data, strips, and scans	12 Tbyte
Merged pixel map	8 Tbyte
Flattened 2D spectroscopic frames	70 GByte
Sky map	500 GByte
Spectra	50 GByte
Atlas images	250 GByte
Parameter list of all objects	25 GByte

Parameter list of objects which have spectra	750 Mbyte

As we noted earlier, there is much exciting science to be done with the data in understanding structures of galaxies and quasars on unprecedently large scales. With 10^6 galaxy and 10^5 QSO red shifts, and 10^8 photometric objects, out to very large red shifts, there are clearly many other opportunities. One example: the data will be ripe for studies of galaxy and cluster evolution to Z=0.5 and evolution of QSOs to Z=5 and perhaps beyond (if any QSOs exist beyond Z=5, a profoundly interesting question).

The data also presents unique opportunities to start exploring the very large scale structure of all matter, beyond galaxies and QSOs. The light from distant quasars is absorbed by the intervening intergalactic gas. The survey will allow a map of the structure of this gas. Beyond that, and probably beyond the capability of the first five years of the Sloan Survey, is the holy grail, a map of the structure of all matter, including dark matter. A promising approach is to use statistical techniques to measure the effect of weak gravitational lensing on the apparent orientation of far galaxies.[3)] From this one can deduce the structure of the gravitational potential, and thereby that of all matter in the observed regions.

Even if this is beyond the seeing capabilities of the SDSS, such a weak lensing survey will have to depend on the kinds of techniques being pioneered by this survey, the automated readout of a dedicated telescope with a large field of view and a focal plane filled with CCDs. One can expect that once proven by the SDSS such automated survey techniques will be applied more and more broadly. Other examples recently discussed are super nova and strong gravitational lens searches. How will the SDSS equipment be used after the first five year survey? We have discussed options which include weak lensing and a move to a southern hemisphere location.

Acknowledgments

Many thanks to Jean & Kim Trân Thanh Vân for their superb organization of this conference in the spirit of bringing the scientific-and the rest of the world-closer together.

I thank my colleagues on this project for their help and support over the years. In particular I would like to thank Rich Kron and Don York for reading this manuscript. The project received a generous grant from the Sloan Foundation which is honored in the name of the survey. Other important funding has come from the Department of Energy (under contract DE-AC02-76CHO3000 with Universities Research Association, Inc., the manager of Fermilab), from the National Science Foundation, the US Navy, the Japanese government through the JPG, and the collaborating universities (Chicago, Johns Hopkins, Princeton, and Washington), and the Institute for Advanced Study, which are members of the Astrophysical Research Consortium (ARC). We are still holding out our tin cup.

Footnotes and References

[1] Geller, M.J., & Huchra, J.P., *Science,* **246,** 897.
[2] Gott, J.R. *et al., Ap. J.*, **340**, 625.

[3] R. D. Blandford, A. B. Saust, T. Brainerd, and J. Villumsen, *Mon. Not. R. Astron. Soc*. **251**, 600 (1991); J. Tyson, F. Valdes, and R. Wenk, *Ap. J.*, **349**, L1 (1990).

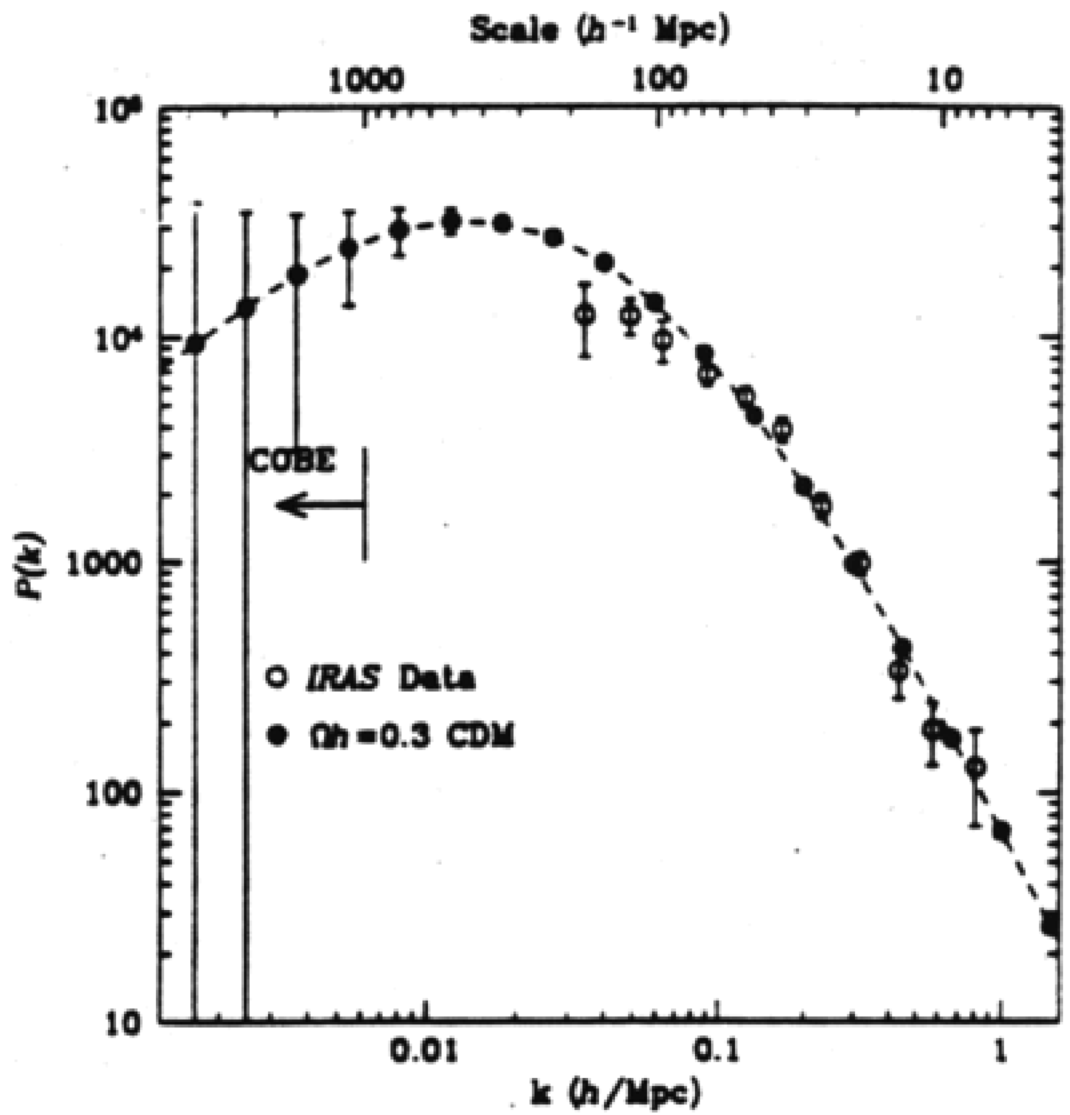

Figure 1. Anticipated SDSS measurement of power spectrum (solid points). IRAS data (open points) shown for comparison. Scales probed by COBE anisotropies are indicated.

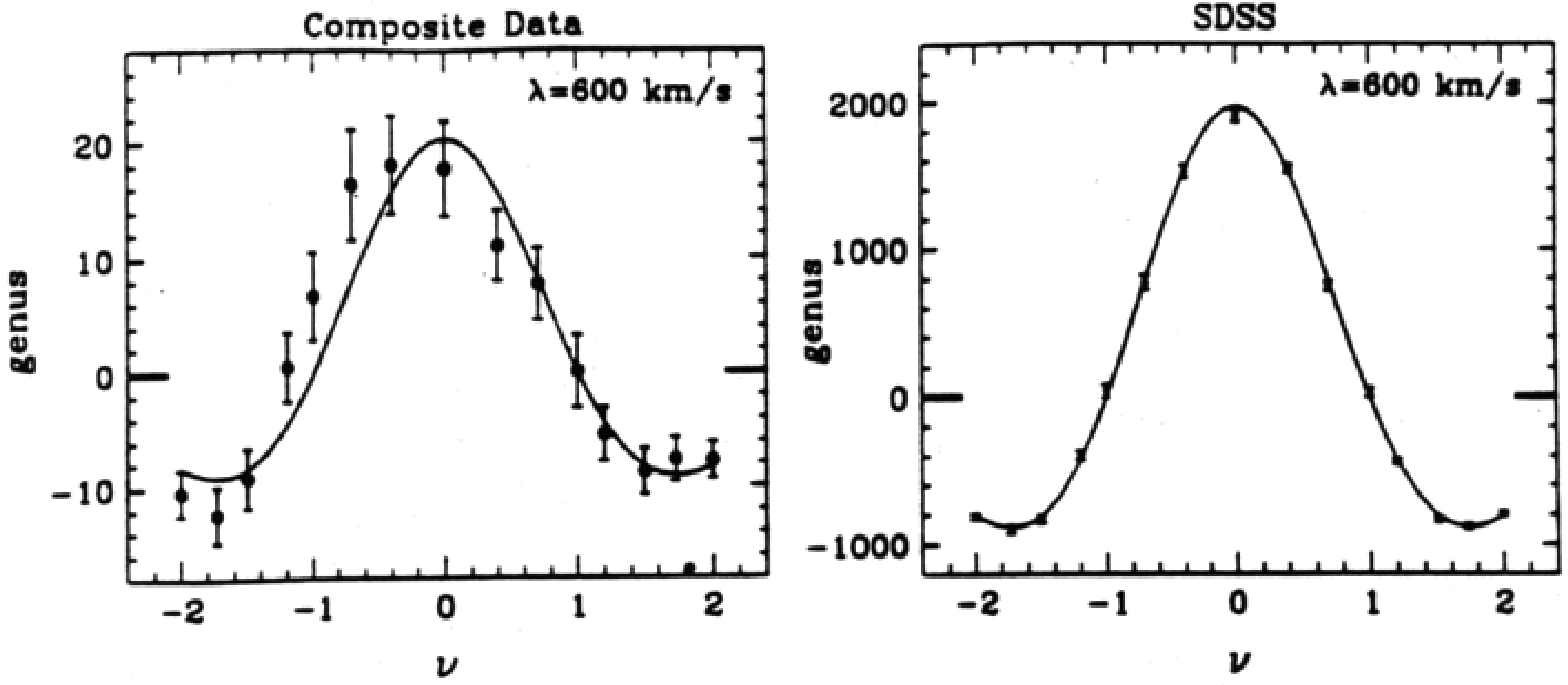

Figure 2. Comparison of quantitative measurement of the topology of galaxy distributions. a) Left, results from existing large red shift surveys. b) Right, results expected from the SDSS. Variables explained in text.

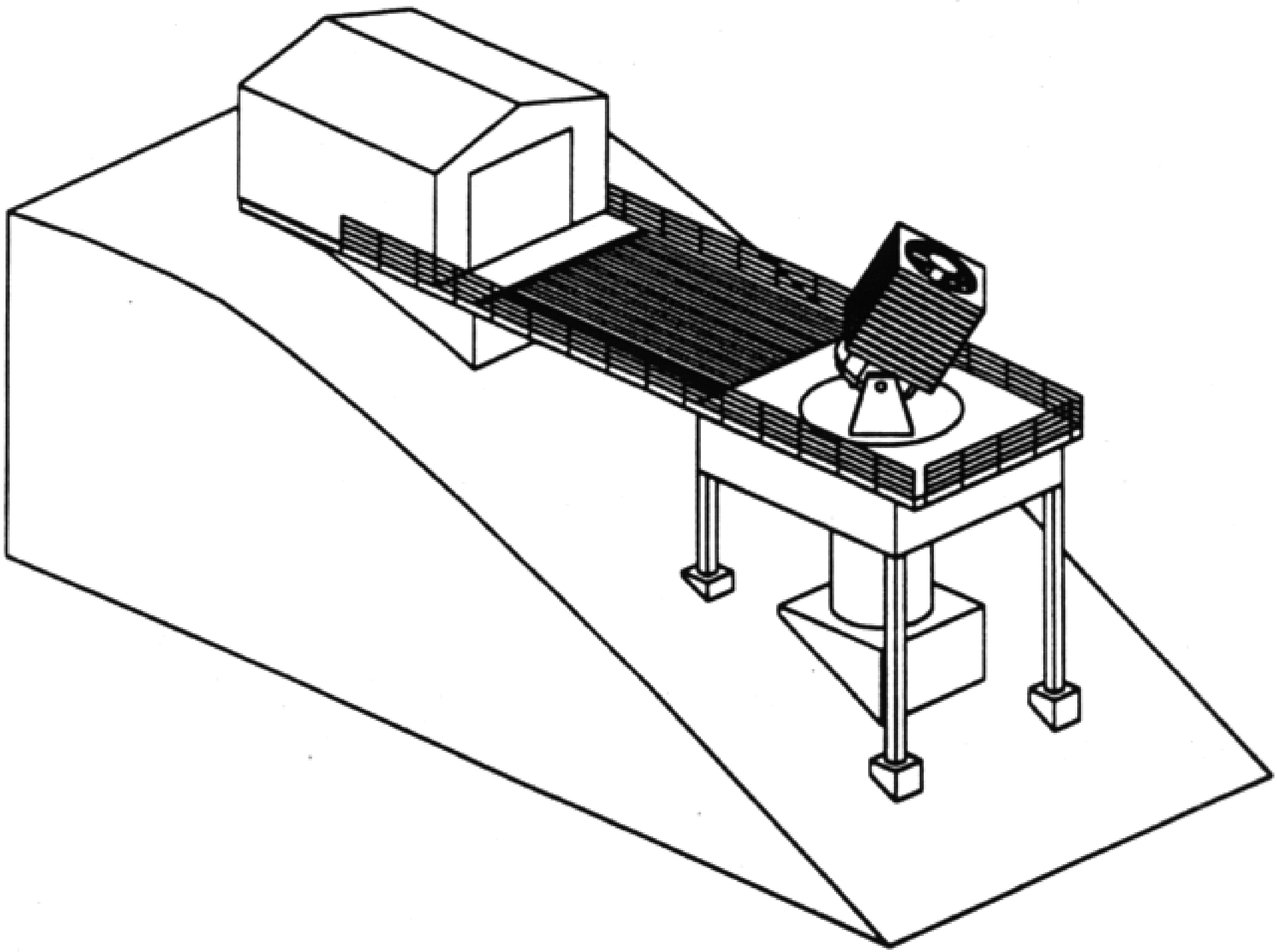

Figure 3. The Sloan Digital Sky Survey Telescope at Apache Point Observatory, NM

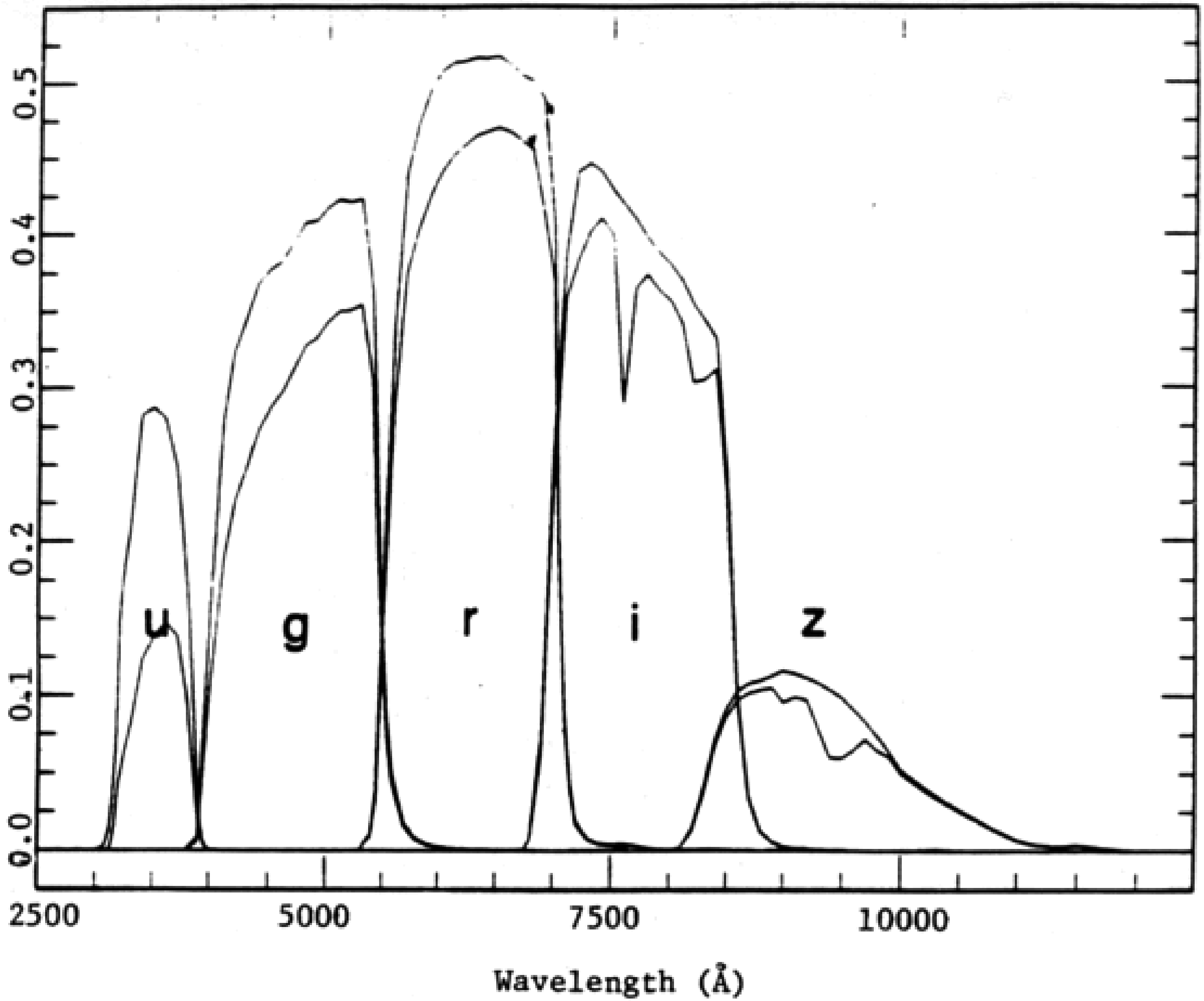

Figure 4. SDSS Filter Response Curves (with and without atmospheric absorption)

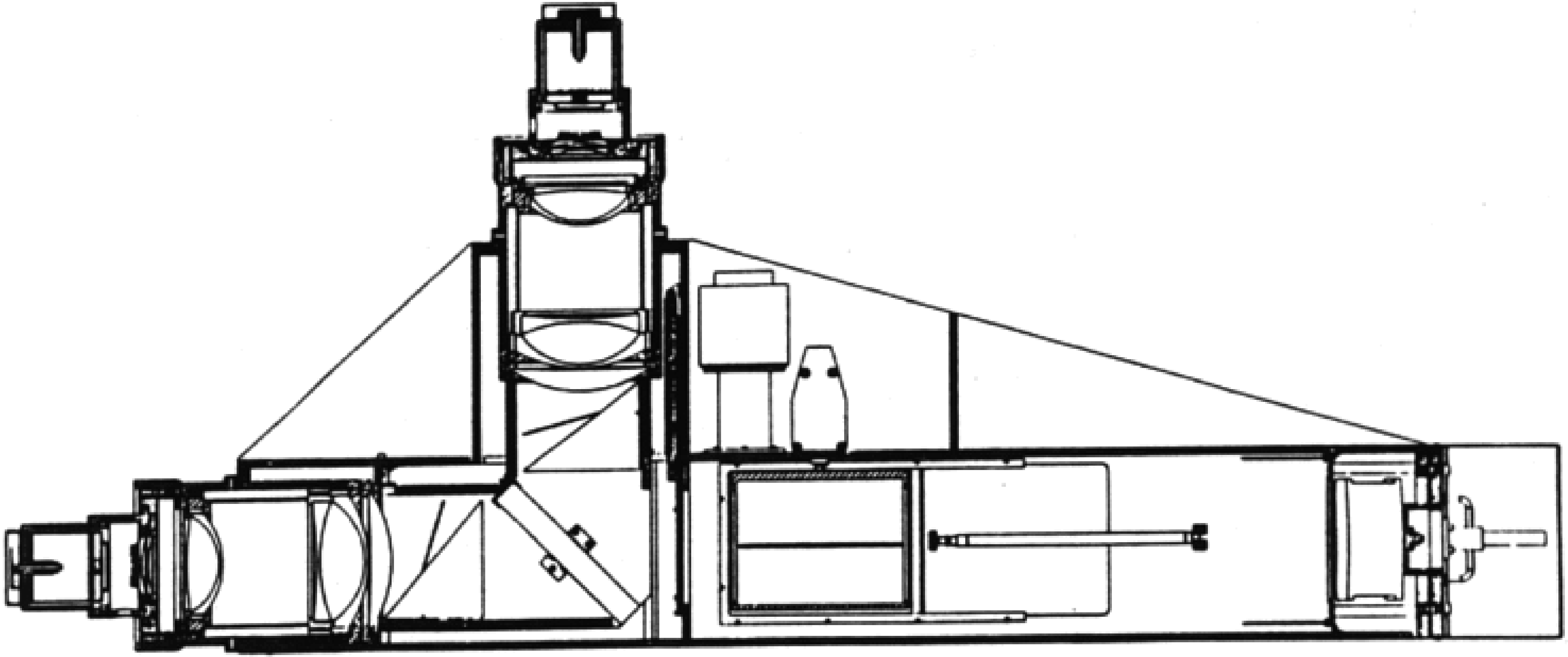

Figure 5. A cross-sectional view of one of the two spectrographs.

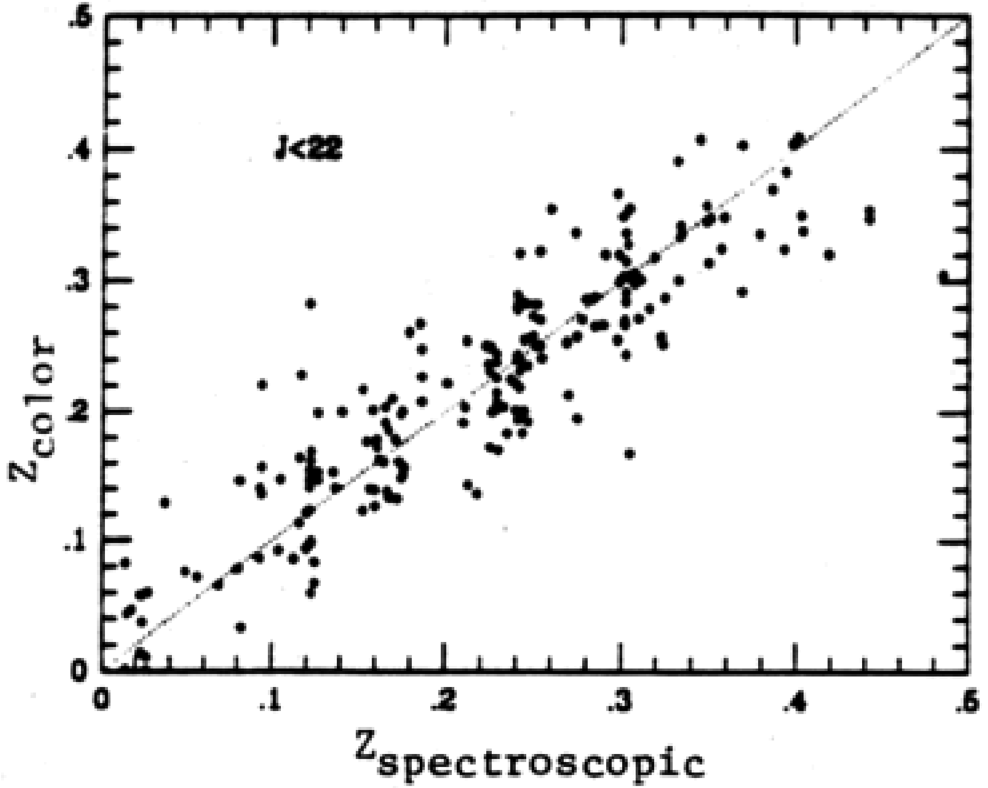

Figure 6. Red shift estimated from photometric colors vs. spectroscopic red shift (from D. Koo & R. Kron, *Ann. Rev. Astron. Astrophys.*, **30**, 613). For the SDSS we expect $\Delta Z < 0.2$.